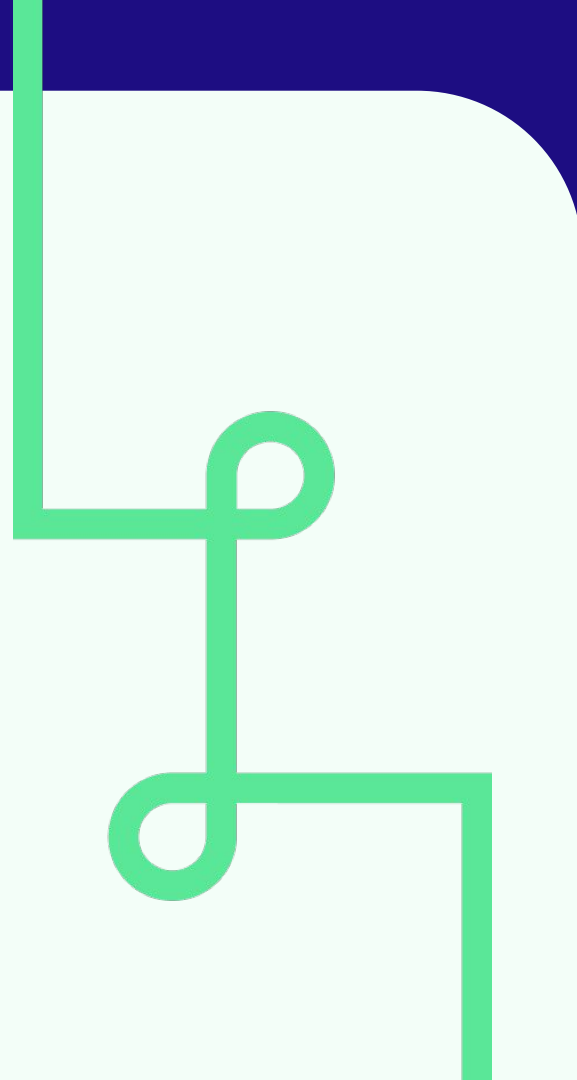




# How Retrieval-Augmented Generation (and AI) can free up time in your organization

Beltug X-change 22/02/2024





## Hi, I'm Frederik

- Enterprise architect
- 20+ years of tech expertise
- 5+ year of enterprise architecture expertise

**Interests:** the digital enterprise -  
technology - sports

## Hi, I'm Steven

- Co-CEO & co-founder Dropsolid
- 20+ years of web & tech expertise
- 10+ years of business expertise

**Interests:** technology - digital experiences - digital strategy





10+ years  
of growth



Constant  
Re-investing



Product &  
Services



€ 8M  
2023  
Revenue



Building  
Partnerships in  
3 continents



Global  
team of 85



€ 3,5M  
self-funded  
platform  
technology



## Enterprise

acerta

beCobank



boost<sup>®</sup> | shopper  
activation



Sleeplife

## Organisations

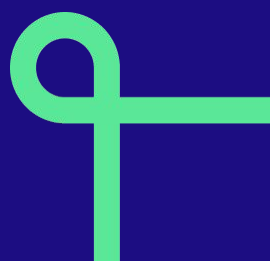



IOS  
Press



## Government





How will you  
leave this session?



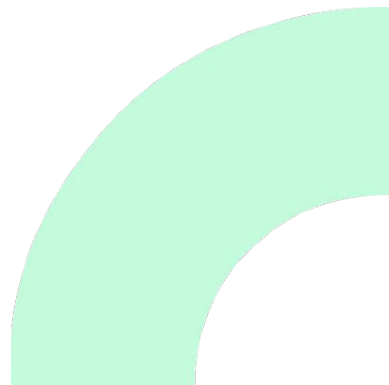
When you see a Magician doing a trick ... at first you don't understand.  
But when the trick is explained ... you feel silly that it's in fact not that magical at all.





# What can RAG & AI mean for your business?

- Lower your TCO by reducing time spent in communication
- Lower your customer acquisition cost by using personalized experiences
- We'll deliver key insights in
  - Marketing automation
  - CRM
  - Website personalization
  - Editor and marketer experience
- And thus overall
  - How to improve your **input/output ratio**
  - How to **gain actual ROI with AI** vs. joining the hype/trend of the moment





First let's look **back**




# from 0 to 100 in drupificial intelligence



Not pizza



✓ Pizza

A photograph showing a group of people in a public space, likely a transit station or airport. In the foreground, a man in a tan shirt is looking at his phone. Behind him, a man in a blue checkered shirt is also looking at his phone. To the right, a woman with curly hair is looking at her phone. The background is blurred, showing other people and structural elements of the building.

By 2020 the average person will have more conversations with their bot than with their spouse.

Microsoft®  
**Research**

# Some ideas

# Automatic img/txt tags

Summary ([Hide summary](#))

AUTOMATIC SUMMARY?

Leave blank to use trimmed value of full text as the summary.

Body

**B** *I* |   |   |   | Format ▾ |  Source

TEST

Text format Basic HTML ▾

[About text formats](#) ?

tags

OR AUTO|TAGS?

# OCR on text uploads

ocr\_file

 no file selected

One file only.

2 MB limit.

Allowed types: txt doc docx pdf.



# Content generation (neural network)

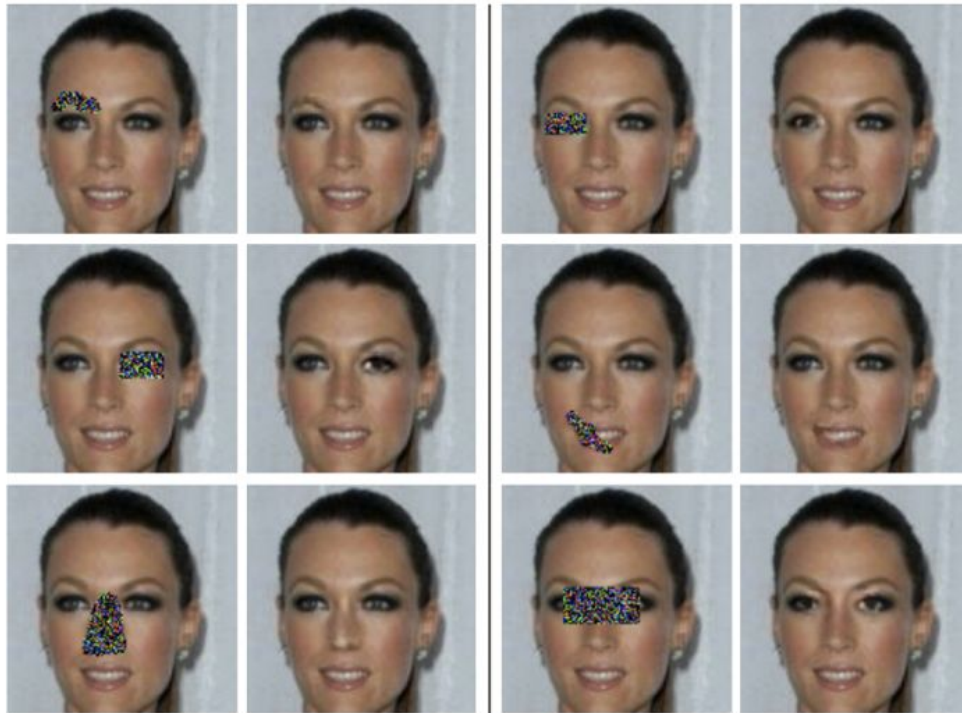


the sun rays struck my face  
warm tingles to my fingertips  
the light showed me a path  
i should walk down  
i spoke and the whispers of the breeze  
told me to close my eyes  
i lost my way in a paradise

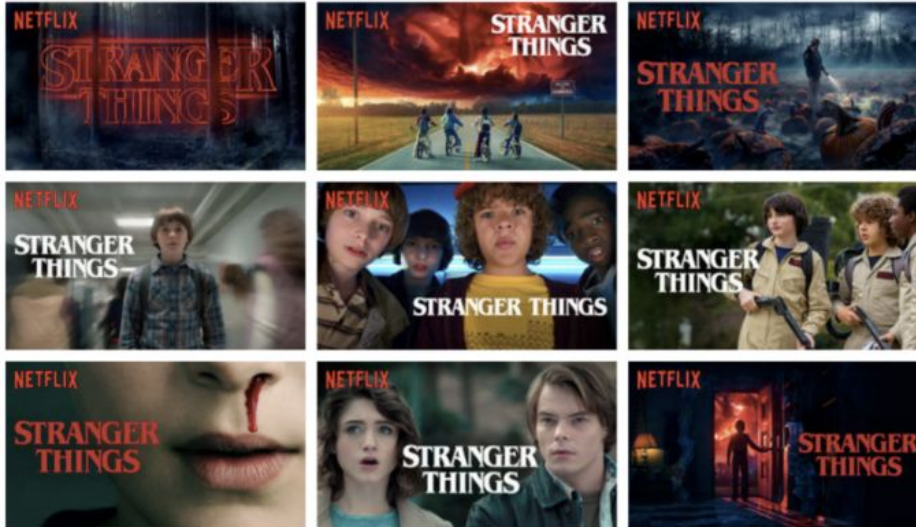


i have been a great city  
spinning and shout  
the sound of the road  
washed away  
the mountain passes through  
the streets are gone  
the silence is raining  
it sits still in silence  
glint its own

# Media improvement (generative adversarial network)



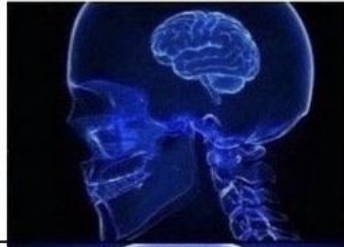
# Content personalisation (a la netflix)



## Suggested content

- [Drupal con London](#)
- [London technology meetup](#)
- [Caching with Boost](#)
- [Creative technology](#)
- [HTML5 video](#)
- [London offices](#)
- [IT sector growth in the UK](#)
- [Best movies 2013](#)
- [Interview with Chuck Norris](#)
- [Review: Independence Day](#)

**HTML**



**CMS ON  
STRUCTURED DATA**



**DYNAMIC  
CONTENT-TYPES  
AND FIELDS**



**AI  
FILLING DRUPAL  
CONTENT**





“



**Frederik Wouters**  
*Enterprise Architect*



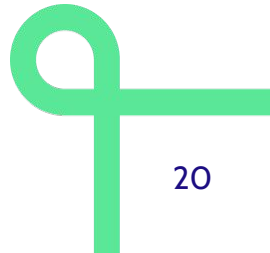


# What is Retrieval-Augmented Generation



## Parts of a RAG

- Embeddings
- Vector Database
- LLM (Large Language Model)







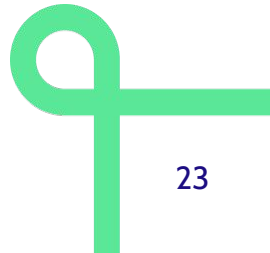
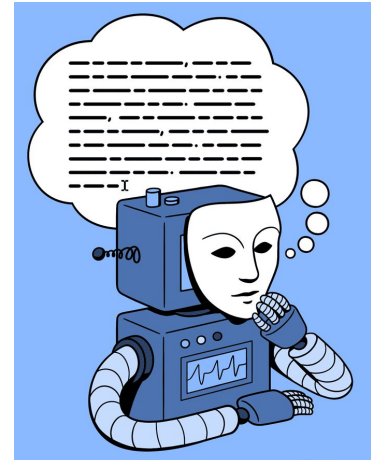


# Large Language Model

*Deep learning algorithm to generate human sounding text*

Important concepts

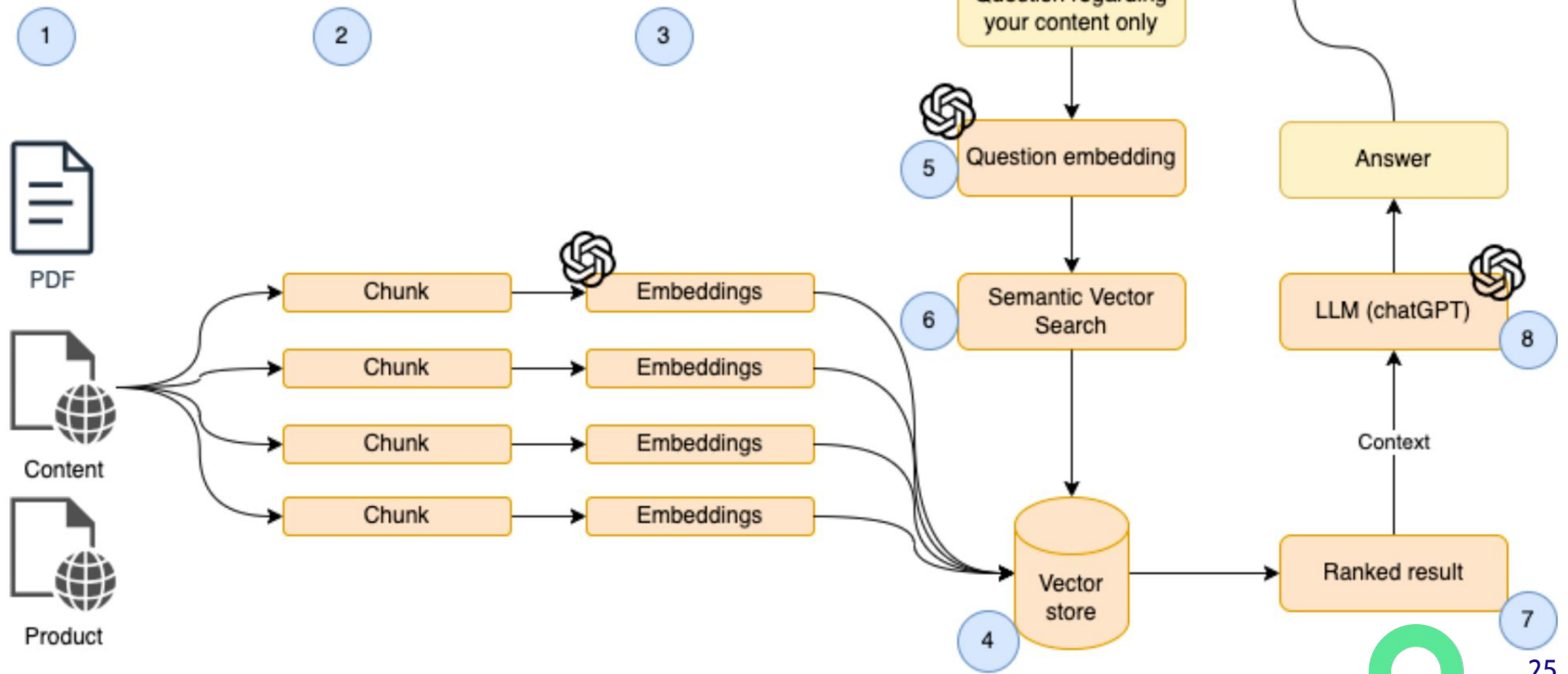
- Prompt
- Tokens
- Temperature





**When these come together:**

**Magic happens**

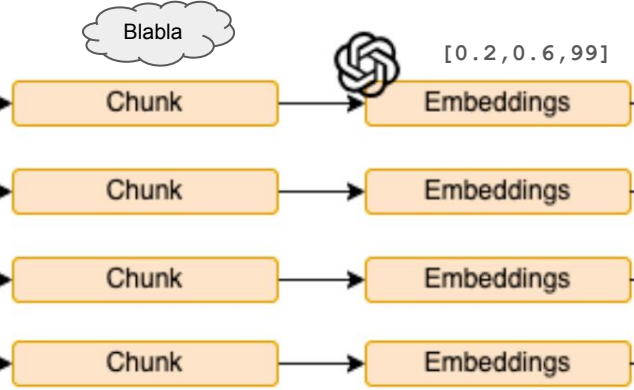




1



2



3



5

Question regarding your content only

How to X?

Question embedding

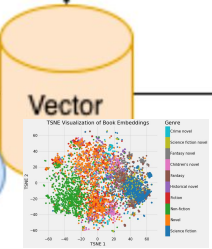
[33, 0.7, 95]

6

Semantic Vector Search

SELECT LIKE [33, 0.7, 95]

4



Vector

7

Ranked result

Context  
ARTICLE1,  
ARTICLE3

LLM (chatGPT)

The answer is X

Answer





LIVE DEMO

# drop solid

EXPERIENCE  
CLOUD



Freeing up time:  
**CRM**



User



Question that is  
answered on the website,  
or in the internal  
knowledge base



Typing reply  
costs **\$\$ TIME**



Reply



Employee



User



Question that is answered on the website, or in the internal knowledge base



CRM



Given our content,  
Given this EMAIL  
Generate a reply



Vector DB



LLM



Validated Reply



Reply

~~Typing reply~~ Validate reply = Fraction of TIME



Employee



Freeing up time:  
**Marketing automation**



Employee



EMAIL TO ALL  
EMAIL TO Segment



Marketing  
automation  
system



User





Employee



EMAIL TO ALL



Marketing automation system



Given the generic EMAIL  
Given this users preferences  
Given our email writing style  
Generate a nice personal EMAIL



Vector DB



LLM



User

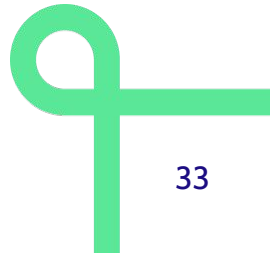


Unique Newsletter  
PER PERSON



Unique Newsletter  
PER PERSON

Unique Emails = Better engagement





# We send out 7.000 unique personalized and fully automated newsletters each week



What is the standard open rate for mails?

**The average email open rate for all industries we analyzed is 21.33%.** Open rates are one of the best ways to tell whether your email strategy is working. This number shows what percentage of your audience opens te emails you send them.

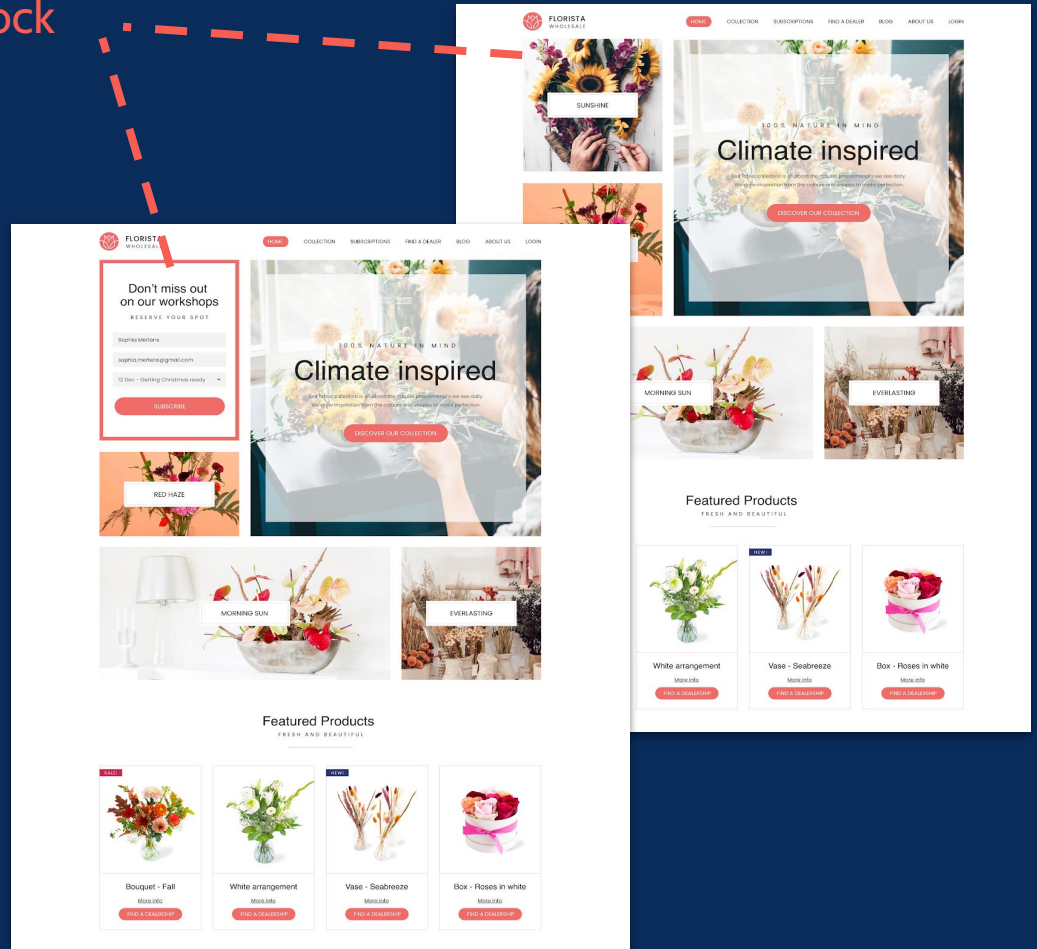
Source: [mailchimp.com](https://mailchimp.com)



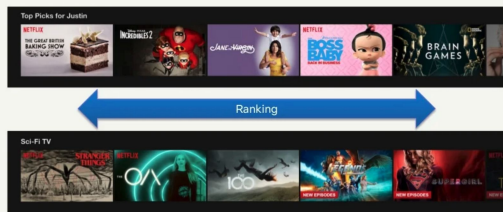
Freeing up time:  
**Website users**

Personalised block

# Personalized homepage

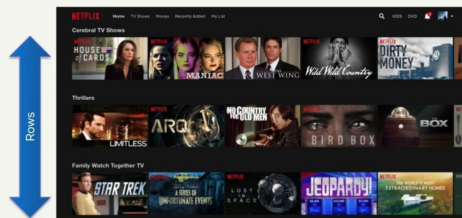


## From what we recommend



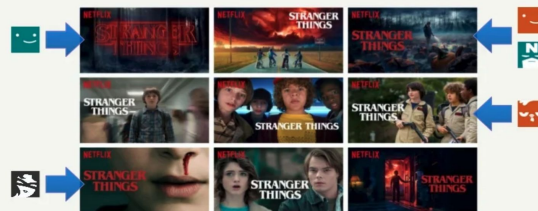
Ordering of videos is personalized

## ... to how we construct a page



Selection and placement of rows is personalized

## ... to what images to select



Personalized images

# Everything is a recommendation!

**Over 80%** of what people watch comes from our recommendations

Overview in [[Gomez-Uribe & Hunt, 2016](#)]

## AI Demo

Based on the provided context from the User's Knowledge Base, I do not have any information about the length of a crocodile.

How long is a crocodile?

Send

## AI Demo

To report phishing attempts, you can use the website [verdachtsafeonweb.be](https://verdachtsafeonweb.be). You can open this website in your email application and send the suspicious email or a screenshot of the suspicious SMS to the provided email address. The service will take care of the rest.

how to report phishing?

Send

## AI Demo

Based on the provided context from the User's Knowledge Base, I do not have any information about the length of a crocodile.

How long is a crocodile?

Send

## Reformatted content from Drupal knowledge base

## AI Demo

To report phishing attempts, you can use the website [verdachtsafeonweb.be](https://verdachtsafeonweb.be). You can open this website in your email application and send the suspicious email or a screenshot of the suspicious SMS to the provided email address. The service will take care of the rest.

how to report phishing?

Send



Freeing up time:  
**Editors**



**Editor**



Write  
CONTENT PIECE  
Facebook POST  
TWITTERX POST  
Linkedin POST  
SUMMARY  
Newsletter summary



**CMS**



Editor

TIME WIN \$\$



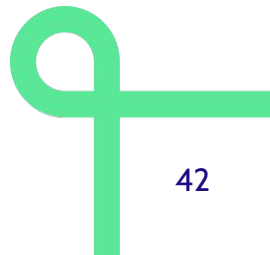
Write  
CONTENT PIECE  
~~Facebook POST~~  
~~TWITTERX POST~~  
~~Linkedin POST~~  
SUMMARY  
~~Newsletter summary~~



Given the content PIECE  
Given our email writing style  
Create social media posts



Facebook POST  
TWITTERX POST  
Linkedin POST  
SUMMARY  
Newsletter summary  
...





Some examples



## Jouw gids voor de overheid

Stel je vraag en vind je weg...

 Vertel, wat wil je weten?



## 🗨️ wat weet je over verhuizen naar nederland?

### 📌 Antwoord

Als u overweegt om naar Nederland te verhuizen, zijn er verschillende zaken die u moet regelen afhankelijk van uw situatie:

#### Nederlandse of EU-nationaliteit:

- Als u de **Nederlandse nationaliteit** of de nationaliteit van een ander EU-land heeft, kunt u zonder verdere voorwaarden naar Nederland verhuizen en hier wonen en werken<sup>[1]</sup>.

#### Britse nationaliteit:

- Voor Britse staatsburgers die niet ook de Nederlandse nationaliteit hebben, is de situatie veranderd sinds het Verenigd Koninkrijk de EU heeft verlaten. Tot en met 31 december 2020 kon men zonder voorwaarden naar Nederland verhuizen. Na deze datum is een **verblijfsdocument** nodig, waarvoor u aan bepaalde voorwaarden moet voldoen. Deze voorwaarden kunt u vinden op de website van de Immigratie- en Naturalisatiedien

Let op! Het antwoord is door een AI gegenereerd. Raadpleeg altijd de onderstaande bronnen.

### 📁 5 handige links

gebaseerd op jouw vraag

 Rijksoverheid

1. Ik wil van het VK terugverhuizen naar Nederland. Wat moet ik regelen?

 Rijksoverheid

2. Wat moet ik regelen als ik in Nederland kom wonen?

 Rijksoverheid

 Rijksoverheid

Vertel, wat wil je weten?

Formatted response

References to CMS

# ZüriCityGPT

Ich bin ZüriCityGPT und ich weiss (fast) alles, was auf [stadt-zuerich.ch](https://stadt-zuerich.ch) publiziert ist.

Frag mich etwas über die Verwaltung der Stadt Zürich.

Frage stellen →



Wir sind weiterhin fleissig daran, ZüriCityGPT zu verbessern, zu erweitern und die Daten regelmässig zu aktualisieren. Wir freuen uns darum über jedes Feedback. Oder gar einem Weitererzählen, falls du zufrieden bist mit ZüriCityGPT.

Und Versuche doch auch die Vorabversion vom mehr chat-artigen ZüriCityGPT: [chat.zuericitygpt.ch](https://chat.zuericitygpt.ch) (noch sehr im Experimentierstadium)

**Hinweise und Einschränkungen**



**Kontakt und Feedback**



# ZüriCityGPT

Ich bin ZüriCityGPT und ich weiss (fast)  
alles, was auf [stadt-zuerich.ch](https://stadt-zuerich.ch) publiziert ist.

Generated Response

References to CMS

Make me a two day city trip, but I'm not very mobile and the weather is bad

Frage stellen →

To plan a two-day city trip in Zurich with limited mobility and bad weather, you can consider the following options based on the provided articles:

Day 1:

- Visit the museums in Zurich to admire ancient and modern art ([Link 1](#)).
- Explore the Technorama or the Umwelt Arena to learn about the laws of nature ([Link 1](#)).
- Experience a guided tour to discover interesting places in the city ([Link 1](#)).

Day 2:

- Enjoy a day at the Alpamare, a sports and leisure center ([Link 1](#)).
- Visit the Zoo Zurich to get up close with animals ([Link 1](#)).
- Attend the "Zwei Tage Zeit" festival for improvised music at Theater Rigiblick ([Link 4](#)).

Please note that these suggestions are based on the available information and may not cover all possible options. It is recommended to check the specific opening hours and accessibility of each location before planning the trip.



But what about my **DATA**





# OS Embeddings

## MTEB or Massive Text Embedding Benchmark

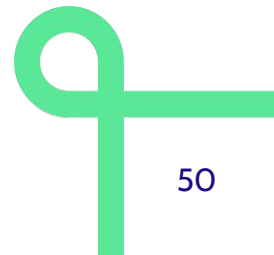
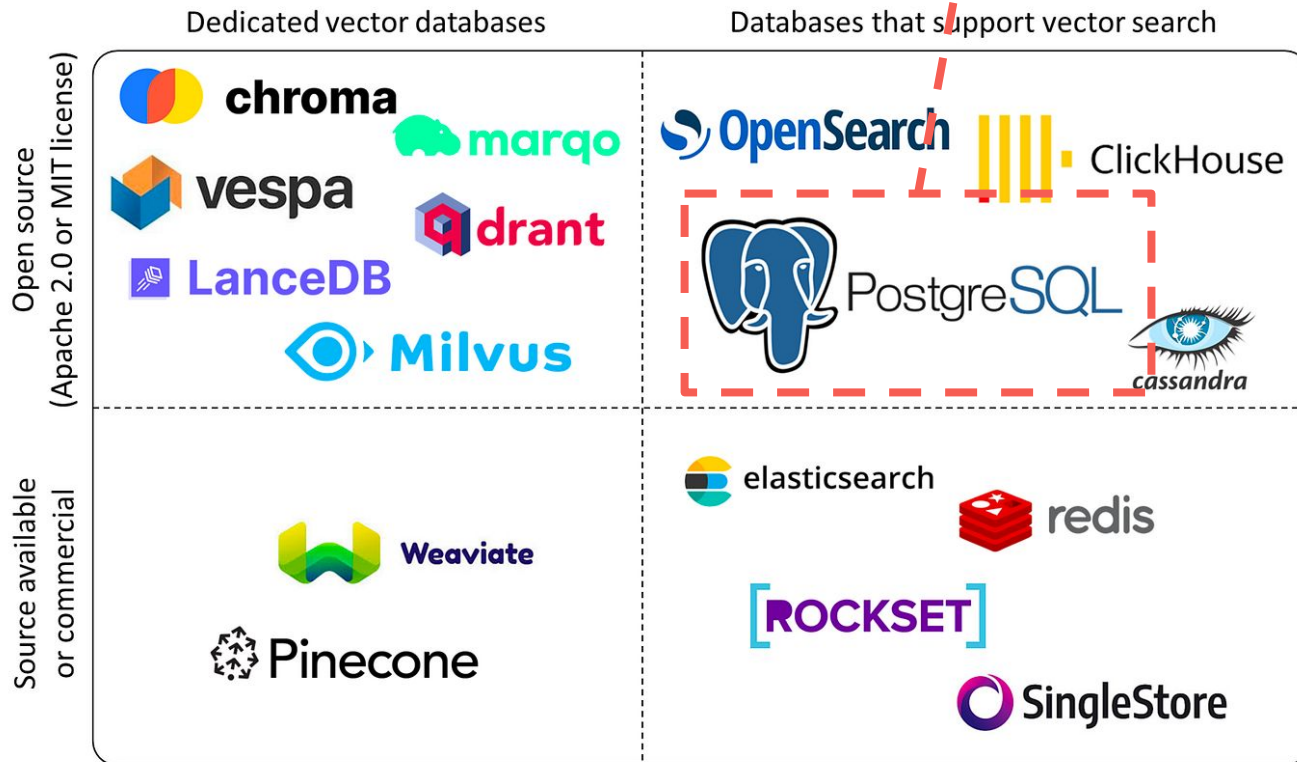
On par with GPT  
WITHOUT THE RISKS

| Rank ▲ | Model ▲  | Average ▲ | ArguAna ▲ | ClimateFEVER ▲ | CQADupstackRetrieval ▲ | DBPedia ▲ | FEVER ▲ | FiQA2018 ▲ | HotpotQA ▲ | MSMARCO ▲ | NFCor ▲ |
|--------|--|-----------|-----------|----------------|------------------------|-----------|---------|------------|------------|-----------|---------|
| 1      | <a href="#">SFR-Embedding-Mistral</a>          | 59        | 67.17     | 36.41          | 46.49                  | 49.06     | 89.35   | 60.4       | 77.02      | 43.41     | 41.88   |
| 2      | <a href="#">GritLM-7B</a>                      | 57.41     | 63.24     | 30.91          | 49.42                  | 46.6      | 82.74   | 59.95      | 79.4       | 41.96     | 40.89   |
| 3      | <a href="#">e5-mistral-7b-instruct</a>         | 56.89     | 61.88     | 38.35          | 42.97                  | 48.89     | 87.84   | 56.59      | 75.72      | 43.06     | 38.62   |
| 4      | <a href="#">voyage-lite-02-instruct</a>        | 56.6      | 70.28     | 31.95          | 46.2                   | 39.79     | 91.35   | 52.51      | 75.51      | 37.93     | 43.7    |
| 5      | <a href="#">voyage-lite-01-instruct</a>        | 55.58     | 58.73     | 37.47          | 45.11                  | 43.42     | 89.71   | 44.79      | 70.46      | 39.66     | 43.33   |
| 6      | <a href="#">text-embedding-3-large</a>         | 55.44     | 58.05     | 30.27          | 47.54                  | 44.76     | 87.94   | 55         | 71.58      | 40.24     | 42.07   |
| 7      | <a href="#">GritLM-8x7B</a>                    | 55.09     | 59.49     | 28.69          | 47.63                  | 46.54     | 85.02   | 49.89      | 73.83      | 35.55     | 39.05   |
| 8      | <a href="#">Cohere-embed-english-v3.0</a>      | 55        | 61.52     | 38.43          | 41.53                  | 43.36     | 88.97   | 42.19      | 70.72      | 42.93     | 38.57   |
| 9      | <a href="#">UAE-Large-V1</a>                   | 54.66     | 66.15     | 38.53          | 40.93                  | 44.89     | 88.24   | 44.84      | 73.13      | 41.4      | 38.65   |
| 10     | <a href="#">bge-large-en-v1.5</a>              | 54.29     | 63.54     | 36.57          | 42.23                  | 44.11     | 87.18   | 45.02      | 74.1       | 42.49     | 38.13   |
| 11     | <a href="#">Cohere-embed-multilingual-v3.0</a> | 53.84     | 55.11     | 29.96          | 40.64                  | 41        | 88.53   | 44.1       | 70.61      | 43.45     | 36.42   |
| 12     | <a href="#">GIST-large-Embedding-v0</a>        | 53.44     | 63.38     | 33.99          | 43.44                  | 42.96     | 86.55   | 44.3       | 70.46      | 41.39     | 38.65   |



# Vector Database

Run it yourself



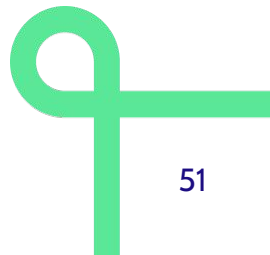


# OS Large Language Model

*Llama 2 is just as accurate as GPT-4 and much faster and more efficient than GPT-3.5 and GPT-4*

## Some examples

- Bert (proprietary)
- Claude (proprietary)
- GPT-3 (Openai = proprietary)
- GPT-3.5 (Openai = proprietary)
- GPT-4 (Openai = proprietary)
- **Llama 2 (Open Source)**
- **Bloom (Open Source)**





Looking to the Future







CONTENT  
Is the  
NEW  
GOD







# Dropsolid - The open Digital Experience Company

Digital experience strategy

UX/UI strategy & design

Enterprise architecture

SEO & digital marketing

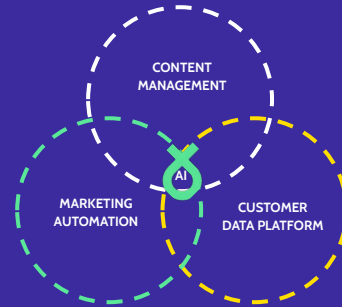
Digital experience integration

Drupal architecture

Drupal integration

Consulting & training

Digital experience platform



Drupal cloud

Digital experience compliance

GDPR expertise

Accessibility

ISMS & Cyber Security